

DOCUMENT RESUME

ED 221 040

FL 013 131

AUTHOR Marshall, Geraldine
TITLE Investigation into the Present State of Language Proficiency Tests and the Possible Use and Development of Communicative Tests for English as a Second Language.
INSTITUTION Coventry Education Authority (England).
PUB DATE 80
NOTE 56p.; Produced by the Minority Group Support Service.
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Cloze Procedure; *Communicative Competence (Languages); Elementary Secondary Education; *English (Second Language); *Language Proficiency; *Language Tests; Linguistic Competence; Linguistic Performance; *Standardized Tests; Test Construction; *Testing Problems; Test Reviews

ABSTRACT

In an appraisal of proficiency tests currently available for English as a second language (ESL), it is found that linguistic competence, rather than communicative proficiency, is being measured. Problems involved in developing standardized communicative tests for use with different age groups in a multicultural school context are discussed. Guidelines are given for designing assessment procedures for communicative proficiency tests. They fall into the general areas of: (1) analyzing the language needs of children in terms of global tasks and communication skills, (2) age relevance and authenticity of materials, and (3) extrapolation and assessment. A summary chart of tests of proficiency in ESL is appended. (EKN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED221040

INVESTIGATION INTO THE PRESENT STATE OF LANGUAGE
PROFICIENCY TESTS AND THE POSSIBLE USE AND
DEVELOPMENT OF COMMUNICATIVE TESTS FOR
ENGLISH AS A SECOND LANGUAGE

By Geraldine Marshall

Geraldine Marshall is a teacher with the Minority Group Support Service which was established in 1977 by Coventry L.E.A. to help meet the needs of ethnic minorities within the local school system.



Minority Group Support Service, Coventry.

COVENTRY EDUCATION AUTHORITY
Director of Education - Mr. R. Aitken

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

D. R. Pritchard

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

FL013131

PREFACE

The possibility of doing some research into the problems of assessing the language proficiency of second language children and examining the state of testing today was first suggested by Coventry L.E.A. in October 1978. Along with this went the idea of writing a report with guide-lines for future assessment procedures.

The suggestion was followed up in November 1978 by four open meetings with teachers in the E2L (English as a Second Language) field to exchange ideas, build up a reference group and cover some of the theoretical background to testing. Finally it was arranged that I would be given secondment for Spring Term 1979 in order to concentrate fully on proficiency tests for children, especially with reference to communicative proficiency. Consequently, I had 13 weeks in which to find out what was available, what were the best theoretical and practical approaches to the problem and finally to write this report incorporating some guide-lines for the assessment of pupils' language proficiency.

The first week was spent drawing up a study/research programme incorporating background reading and a plan for a series of visits to people who would probably be able to help.

From this week, a number of things emerged: a time-schedule for the whole project; a bibliography and collection of books and articles on the subject, (greatly added to during the project) and a list of known tests available and possible sources of others.

The meetings with local teachers produced the following key questions which we felt needed to be answered.

The present state of testing in E2L

- a) What kinds of E2L proficiency tests are most appropriate?
- b) What exactly is meant by 'communicative proficiency'?
- c) What materials have been or are presently being developed for E2L proficiency testing?
- d) What relevance do these materials have for school children (given age range 4 to 17 years)?

Theoretical and practical problems of testing

- a) What are the main features of communicative tests and how are these tests developed?
- b) What are the problems which need to be solved in the development of communicative tests?
- c) What is the difference between communicative competence and communicative performance?
- d) What do integrated tests like Cloze tests and dictation tests measure and how useful are they?

Relationship of above to E2L in the British Educational System

- a) What do we know about the natural language developments of mother tongue children and their expected levels of proficiency relative to age?
- b) Do different ages need a different approach, or just different materials?
- c) Which would be the best ages to group together?
- d) What language banks are available that are specifically related to the language needs of children?
- e) How are language banks and test batteries developed?
- f) Which types of items and formats are most suitable for children?

INTRODUCTION

The aim of this work is to examine the present state of proficiency tests* in the E2L field and to look at the advantage and possibilities of developing communicative assessment procedures.

This report is presented as follows:

- PART 1: General and Theoretical Appraisal of Testing English as a Second Language. (Pages 2 - 14)
- PART 2: The Research Programme. (Pages 15 - 21)
- PART 3: Addressing the research findings to the practical problems of E2L teaching in multi-racial schools. (Pages 22 - 31)
- PART 4: Guidelines for drawing up assessment procedures for communicative proficiency tests. (Pages 32 - 36)
- PART 5: Conclusion and Recommendations. (Pages 37 - 38)
- Appendices I - III (Pages 39 - 42)
- Glossary (Pages 43 - 46)
- Bibliography (Pages 47 - 49)
- Acknowledgements (Pages 50 - 51)

During my research programme, my energies were directed towards three tasks:

1. To find out exactly what is the present state of testing and what communicative testing materials are already available and suitable for school children in multi-racial British schools (Part 1).
2. To explore both the theoretical and practical problems which need to be solved if such tests are to be developed (Parts 2 and 3).
3. To relate these elements to the British educational system (Part 3).

* See glossary for this and other terminology that occur in the text.

PART 1
GENERAL AND THEORETICAL APPRAISAL
OF E2L TESTING

- A. THE PRESENT STATE OF TESTS AND ASSESSMENT PROCEDURES RELATING TO E2L.
- i) Introduction
 - ii) The published E2L tests available
- a critique -

Pinpoint

N.F.E.R. Tests of
Proficiency in English

C.S.E. Modes I, II and III
and University Boards
 - iii) Alternative approaches - checklists
- B. BRIEF REVIEW AND CRITIQUE OF CERTAIN APPROACHES TO E2L/EFL TEACHING AND TESTING (with reference to the Morrow (1978)⁽¹⁾ and Davies (1978)⁽²⁾ articles.
- i) Basic changes in approaches to language teaching and testing.
 - ii) The influence of these changes in relation to E2L testing.
- C. SUMMARY

-
1. MORROW, Keith - 1978
Communicative Language Testing: Revolution or Evolution. (Will appear in 'Communicative Approach to Language Teaching. Editor K. Johnston, OUP 1979.)
 2. DAVIES, Alan - 1978
(July & Oct. 1978) Language Testing: Survey article (Language Teaching Linguistics: Abstracts Vol. II, No. 3 & 4, C.U.P.).

A. THE PRESENT STATE OF TESTS AND ASSESSMENT
PROCEDURES RELATING TO E2L.

i) Introduction

If teachers are to be truly effective in the field of English as a Second Language (E2L), it is clear that they need to draw up some kind of specific teaching programme according to the needs of their pupils. Consequently we need to know in fairly precise terms, exactly what stage our pupils are at. In order to assess this, many teachers feel the need for some form of standardised test material.

The first type of test required is a test of proficiency, i.e. assessing the learner's ability to use English at the time of testing.

Secondly, a diagnostic test or procedure is necessary to find out precisely which of the communicative skills and elements of language performance the child has mastered or which he still needs to master. This kind of knowledge enables the teacher to prepare future teaching programmes.

In Coventry, a number of our minority group children are given E2L help either within the classroom or in withdrawal groups. The decisions as to which children need to be given special help and for how long depend on a number of factors besides the children's ability in English. These factors include for example, their degree of self-confidence and the extent to which they have overcome any cultural shock or disorientation, etc. Performance in E2L cannot be isolated from these other social and psychological factors. Assessment of language ability must be handled sensitively and with an understanding of the personal difficulties under which learning may be taking place. However success in English can be an important aid to an individual's adjustment and development. There is a strong case then for effective E2L teaching and careful monitoring of pupil's progress.

Proficiency tests can help to provide useful information by measuring a child's progress and identifying children for whom language help should be a high priority. Proficiency tests have already helped to reaffirm the need for specialised E2L teaching by providing strong evidence of linguistic underachievement in some children.

ii) The Published E2L tests available - a critique -

Given the stated need for standardised objective assessment procedures it is a sad fact that any survey of the field of English as a Second Language immediately shows up the great lack of commercially published materials, both for teaching purposes and for the assessment of children. This is particularly true when the amount of commercially produced E2L materials is compared with its more 'lucrative' sister subject: English as a Foreign Language (EFL).

In December 1976 the British Council published an Information Guide (E.T.I.C. No. 6) Called 'Examinations and Tests for Speakers of Other Languages'. The tests listed were mainly proficiency tests. However, only 2 out of the 20 listed are for E2L purposes. These are:

- i) 'Test of Proficiency in English' (N.F.E.R.)
- ii) 'Test of English as a Second Language' (Georgetown University, U.S.A.)

Only the former relates to the English education system.

Although there are a number of E.F.L. proficiency tests, which may have value at the 16+ stage, they cannot really be used by E2L teachers, for two basic reasons. First, they lack face validity, i.e. they are not culturally relevant for many of the children. Secondly, they are often inappropriate for the full age range of children because, on the whole, E.F.L. proficiency tests have been written for roughly one age-level of proficiency, i.e. either for school-leavers or students wishing to enter higher education. However, it is clear that for E2L school children, several levels of proficiency tests are required reflecting the different levels of skills, concepts and interest relative to the children's ages. For instance no-one would expect reading or writing proficiency assessment from a 4 year old infant.

Of the few tests that have been published, the best known are probably as follows:

Pinpoint

The Pinpoint test was designed by Patricia Cannon and colleagues of the Leicester Language Centre. It follows a checklist approach and has been useful for diagnostic purposes and in helping teachers who are conscious of the importance of methodical language development for every child. It is a language framework for use in nursery and infant classes and it has 3 main parts:

- (a) The diagnostic test which is set up by the teacher to suit her own situation. Each section of the test deals with a different

language area and uses an individual 'checklist' record sheet. Objects from the classroom and a book of special photographs illustrate a variety of school-based actions and situations.

(b) Seven 'Language Check Cards' which set out the fundamental core of language to be covered by the end of the infant period.

Check Card 1 - 'Instructions'

Check Card 2 - 'Particular Formulae
and Prepositions'

Check Card 3 - 'Sets of structural words'

Check Card 4 - 'Adjuncts'

Check Card 5 - 'Basic sentence patterns'

Check Card 6 - 'Verb and Noun patterns'

Check Card 7 - 'Tense forms of verbs
with extension by word,
phrase and clause'

(c) Planning Notes for setting up special language teaching situations and ideas for exploiting existing ones.

National Foundation for Educational Research in
England and Wales (N.F.E.R.)
Tests of Proficiency in English, (September 1973)

The tests were drawn up for classroom use and were specifically designed as E2L tests for junior schools. They are intended to have a three-fold function:

Firstly, these tests set out to help with decisions about whether children should attend special language classes and if so when they should return to 'normal' classes.

Secondly they can be useful as a diagnostic guide for future teaching. By analysing the children's errors the teacher can be helped to formulate E2L teaching programmes.

Thirdly they can be used in monitoring the progress of E2L children, individually or collectively.

As with 'Pinpoint' N.F.E.R. tests are basically content-reference i.e., the scores indicate the child's performance in terms of the skills tested and do not rank the pupil in relation to other pupils or to language competence as a whole.

The framework within which the test materials have been constructed is a structural linguistic one. Consequently, although these tests may be of use for diagnostic purposes to analyse a pupil's formal linguistic competence, they

are not appropriate for assessing the communicative proficiency of a pupil because they do not measure language in action. A Table of Contents of the tests and skills identified in the N.F.E.R. tests is shown in Appendix 3. The table illustrates a purely taxonomic approach which is unsatisfactory in so far as it cannot measure a child's ability to use language in real-life situations. The taxonomic approach involves the categorisation of language into various levels and elements, such as phonology, grammar semantics and vocabulary. Phonology, for example can be sub-divided into: phonemes, stress, intonation and rhythm. Such an analysis is in terms of structural linguistics. It stresses elements, not communication.

There are other possible criticisms of the N.F.E.R. tests. Firstly illustrations may have a cultural loading e.g. the chef or the cake.



Secondly, some pictures are not always clear as to the point they are illustrating. This is because the way pictures are perceived is partly culturally influenced and is also dependent on a child's age. e.g. in the pictures below many children with whom I conducted the test chose picture D to illustrate 'She is buying some apples'. This was based on the perception that women usually buy apples and that there was nothing strange in a boy working in a shop. In fact, the boy was often seen to be offering change.



The woman is buying some apples

Reproduced with permission.

C.S.E. Modes I, II and III and University Boards

In recent years another small group of proficiency tests has been developed mainly through pressure on Regional Examination Boards from individual secondary schools. This pressure grew as teachers saw that without some recognised certification of attainment in English, second-language school-leavers were at a disadvantage when seeking employment or entry to further education. Some very able minority group children have been held back by failure in English 'O' level examinations. Present 'O' level English language papers are a poor indication of a child's overall ability to learn through the medium of English. This is shown by the number of pupils getting high grades in other subjects but failing English 'O' level.

It was also clear that the prescribed syllabuses and examinations for indigenous candidates were unsuitable for examining many E2L learners because of their degree of figurative content and specific cultural bias. Only 2 out of the 14 C.S.E. examination boards have provided for E2L learners by preparing proficiency tests in E2L Mode I level. This is despite the fact that within this specific age range and level the language needs of pupils can be more clearly defined.

The two Examining Boards with Mode I E2L exams are:

- (1) West Midlands E.B.
- (2) North West Regional E.B.

So far, I know of no Mode II E2L exams and there are only three boards who have C.S.E. Mode III E2L syllabuses and exams. These are:

- (1) Associated Lancashire Schools E.B.
One Mode III for a 'number' of Bolton schools.
- (2) East Anglia E.B.
Two schools with Mode III.
- (3) East Midlands Regional E.B.
One school only in Derby with Mode III.

-
1. By 'figurative' I mean language and words used not in their literal sense but in an imaginative way, e.g. when 'fiery' is used of a man who is easily made angry.
 2. So far, replies have only been received from 10 out of the 14 C.S.E. Examining Boards so there may possibly be other Mode I and III C.S.E. E2L Exams.

There is no G.C.E. 'O' level E2L examination but there are two other bodies who set E2L/EFL proficiency examinations for entrance into further higher education. These are:

- (1) North West Regional Advisory Council for Further Education.
- (2) Northern Universities Joint Matriculation Board.

There is a particular problem with C.S.E. exams in E2L*. Although it might be thought that a school leaver with an E2L English certificate is preferable to a pupil with no qualifications in English, in times of high unemployment when employers can afford to be choosy, E2L English is often seen as no substitute for C.S.E. English or G.C.E. 'O' level English.

The general level of achievement required in these C.S.E. exams is established by comparing performance with what is appropriate to the language level of the average 16 year old in Britain today. Consequently they are only suitable for that age group. The tests do partly attempt to assess how proficient pupils are in using language but the general approach is not a communicative one. The tests are often a mixture of discrete-point, structuralist items such as multiple-choice grammar sections and integrated-skills items such as Cloze tests.

iii) Alternative approaches - checklists

As so little material has been available to help teachers to assess English proficiency (N.F.E.R. Tests apart), practising E2L teachers have begun to build up their own assessment methods. These non-standardised 'tests' often take the form of a checklist of structures. The choice of structures is often made on an ad-hoc intuitive basis derived from the teachers' practical experiences of the 'survival language' pupils need in school. These checklists have basically been useful for structural diagnostic purposes but they are unable to accurately measure the true proficiency of a child in communicating and using English language. This is because language is dynamic and so in 'language use' the whole is bigger than the parts; "Knowledge of the elements of a language is not equivalent to knowledge of the language, and in fact counts for nothing unless the user is able to combine the elements in new and appropriate ways to meet the linguistic demands of the situation in which the learner wishes to use the language" (Morrow), i.e. a child must have communicative competence and performance. It is a learner's proficiency in this sense that we should be assessing.

*An interesting example of a C.S.E. Mode I E2L paper is that offered by the N.W. Regional Board. This is described more fully in part 3.

Apart from the theoretical criticisms mentioned above there are practical problems with checklists even when used for diagnostic purposes. The choice of items is usually arbitrary. If the lists are short enough *to be used,* *they will* to be too superficial and if they are detailed enough to be sufficiently informative, they become impractical to administer and score.

B. BRIEF REVIEW AND CRITIQUE OF CERTAIN APPROACHES TO E2L/EFL TEACHING AND TESTING (with reference to the Morrow (1978) and Davies (1978) articles)

i) Basic changes in the approach to language teaching and testing

The beliefs we hold either consciously or subconsciously about the nature of language and how a language is learnt affect our approach to language teaching and testing. They help to determine not only how we teach but how we assess our pupils.

From the mid 1960's to the mid 1970's there were major developments and changes of emphasis in the explanation of language behaviour, (of Chomsky, Halliday and others).

During this period there was a growing interest in language as a means of inter-personal communication and as a social institution. These developments began to show up the narrowness and the apparent inadequacy of a purely 'formal' linguistic description of language which only relates to a language user's 'grammatical' competence. It is clear that with the re-emphasis on communication, the development and assessment of a learner's communicative competence is also important. We might take the example of an E2L student who had 'learnt' a certain range of structures but was unable to use such structures effectively in communication with others.

In terms of language teaching there has been a move away from a purely structuralist approach towards a more functional one. Experimentation with Notional Syllabuses* has entailed concentration on meaning, contextualisation, authentic materials and on the communicative functions of language. This is now being reflected in the new approaches to language testing. The overall change is characterised by A. Davies (1978) "as a move from -

linguistics	to	sociolinguistics
structuralism	to	functionalism
taxonomy and breaking		integration and
into skills and into		building up into
discrete parts	to	wholes"

To give a slightly wider perspective and to develop the discussion we can use the 3 stages that Spolsky (1975) has identified in a recent history of language testing.

*particularly weak forms of Notional Syllabuses
(see glossary)

Stage 1: The pre-scientific stage

Tests of this kind consist largely of translation, composition, and sentence completion. They take no account of statistical validity and reliability which characterise stage 2 tests.

Stage 2: The psychometric - structuralist stage

One of the main exponents of the kind of tests at this stage (between 60's and mid-70's) is R. Lado (1961). The psychometric structuralist approach towards testing was particularly associated with the structuralists and with structural linguistics. This approach has the implicit assumption that mastery of a language is possible by the isolated mastery of its separate parts. By relating the tests of this stage to a particular language theory they were attempting to achieve validity (i.e. construct validity) for the tests. If the basis of the structuralists theory is questioned as it has been by the people like D. Hymes and D. Wilkins (1976), the validity of the tests themselves become questionable.

The concept of validity is circular. This is because validity exists only in terms of its own specified criteria and assumptions. For instance, if a test is constructed on the basis of structural linguistics and behaviourist learning theory the test could be valid but only as a structuralist type of test and not as a communicative one. In other words a structural test can have validity in its own terms without measuring a pupil's communicative performance. This can happen in a number of ways. Firstly, the construction and content of a syllabus may be based on a language learning theory not related to the aim of communicative competence. Secondly validation of the test may be set against another test not concerned with this type of competence. It is seen then that if the assumptions or criteria on which validity is based are false then the validity is also spurious.

The structuralist theory also influences what is to be tested and how it is to be tested. Discrete items such as verbs, adjectives etc. are selected and tested e.g. by multiple choice. Each is intended to reveal the candidate's ability to handle one item of language in terms of each of the four skills: listening, speaking, reading and writing. However, it has proved difficult to test items through any one of the four skills in isolation.

Stage 2 tests also claim to have more reliability through greater objectivity, but this objectivity is only in terms of actual marking assessment. In terms of the evaluation of the numerical score yielded and perhaps more importantly, in terms of the construction of the test itself subjective factors play a large part, (Morrow 1978), i.e. in the selection, type, order, difficulty, number and the precise forms of the items chosen.

Criticism can also be made of the behaviourist language-learning theory connected with the structuralist approach. An alternative view of language learning would hold that the answers to tests can and should be considered as more than just right or wrong. It is not the number of items that are correct but the quality of language performance which is important and it is unsafe to assume that the score on correct (discrete) items allows conclusions to be drawn about the quality of language performance (Morrow 1978). A low score may not necessarily predict ineffective communication.

The acquisition of a new language is always a developmentally creative process. One reason underlying the idea that item scores do not necessarily measure language ability is the view that learners possess 'transitional competence' (Corder 1975) which is dynamic. This produces 'inter-language' (Selinker 1972) which, with further learning, will move towards an approximation of a native-speaker system of language. The role of testing then becomes one of showing how far the learner's competence has moved in this direction, i.e. the test is "thus concerned with making a learner produce samples of his own inter-language based on his own norms of language production so that conclusions can be drawn from it. Tests of receptive skills would be similarly concerned with revealing the extent that a child's processing abilities match those of native speakers". (Morrow 1978)

Present language tests fail to measure communicative competence and performance because (as Morrow points out) there are certain characteristics of communication which are not measured in the conventional existing language tests. These are that communication:

- (1) is inter-action based
- (2) is unpredictable
- (3) has context
- (4) is purposive
- (5) the language is authentic and behaviour-based

i.e. What is said by a speaker depends crucially on what is said to him in a particular context and at a particular time.

Thus, given the increasing importance placed on assessing the communicative use of language and the inadequacies of stage-two types test to do so, the pressure is growing for a communicative approach to language testing to assess overall language ability.

Stage 3: The psycholinguistic - sociolinguistic stage

The key feature of these tests is that they can take an integrative or global view. In such tests whole or 'real' tasks of communication are presented to the students to assess their proficiency both in production and in comprehension.

As yet, there are no standardised published tests for this stage. However, experimental ones are being developed and the Royal Society of Arts already has a pilot test for assessing the communicative proficiency of E.F.L. students.

ii) The influence of these changes in relation to E2L testing

Firstly, testing must take as its starting point the measurement of what the pupil can actually achieve through language. Tests therefore need to have a format which enables a child to produce his own language and the test must not only measure the degree of discrepancy between the language of the child and the language expectations of the assessor.

Secondly, having already noted the features of communication which conventional 'Stage 2' type tests fail to measure, Keith Morrow lists the following characteristics which can be expected of any test of communicative ability.

- (1) It will set out to show whether, or how well, the candidate can perform a set of specified activities. i.e. it will be criterion-referenced against the operational performance of a set of authentic language tasks.
- (2) It will be crucially concerned to establish its own validity as a measure of those operations it claims to measure. Thus content construct and predictive validity will be important, but concurrent validity with existing tests will not necessarily be significant.
- (3) It will rely on modes of assessment which are not directly quantitative, but which are instead qualitative. It may be possible or necessary to convert these into numerical scores, but the process is an indirect one and recognised as such.
- (4) Reliability while clearly important, will be subordinate to face validity. 'Objectivity' will no longer be a prime consideration, although it is recognised that in certain situations test formats which can be assessed mechanically will be advantageous. However, the limitations of such formats will be clearly spelt out.

C. SUMMARY

During the last 15 years, developments in the theory of language and language learning have emphasised language as a means of communication. In spite of this, no tests of communicative proficiency are available in E2L. Proficiency tests based on the structuralist approach, measure linguistic competence rather than communicative proficiency. The nearest to this approach is the N.W. Regional C.S.E. Board's 'English As A Second Language'. Given the need to assess communicative proficiency, we consequently need standardised communicative tests. A few are being developed in E.F.L. but these are not appropriate for all age ranges, only for school leavers. They are also inappropriate in the multi-cultural context of the English educational system and therefore cannot be used as E2L tests. What seems to be needed is research into the feasibility of developing E2L communicative tests for different age-groups of children in the multi-cultural school context.

PART 2

RESEARCH PROGRAMME

- A. 'LANGUAGE TESTING' (Alan Davies)
- B. RESOLVING CERTAIN DIFFICULTIES IN THE
DEVELOPMENT OF COMMUNICATIVE TESTS
 - i) Extrapolation
 - ii) Assessment
- C. CLOZE TESTS AND DICTATION
- D. SUMMARY

A. 'LANGUAGE TESTING' (A. Davies)

For a general and fairly exhaustive survey of testing and the general developments within this field, Alan Davies' article 'Language Testing'* proved to be of immense value. Discussions with Davies and A. Howatt in Edinburgh, helped me to explore further the theoretical background to testing.

The first part of the Davies article discusses discrete point tests with particular reference to the major text books on language testing. This is followed by a short discussion of reading comprehension tests which in turn is followed by a note on criterion-reference testing. The second part of the article discusses integrative tests with special consideration of Cloze procedure and dictation. Next it deals with problems involved in tests of spoken and written production and concludes by looking at the testing of communicative competence.

However, Davies tends to minimise the difference between the psychometric-structuralist approach and the communicative approach. In his discussion of integrated versus discrete point testing, Davies argues that they are at different ends of the same continuum rather than in different spheres.

Davies believes that it is very difficult if not impossible to construct a scorable standardised communicative competence test based on an integrative design i.e. where the parts of the language are integrated and tested in combination with one another. He also concludes that "a convincing argument in favour of linguistic competence tests both discrete point and integrative is that grammar is the core of language learning..... (and) grammar is far more powerful in terms of generalisability than any other language feature." (Davies 1978, p225)

However Davies notes disagreement on this point and cites Savignnon whose study of learners of French concludes that valid communicative tests can be designed and made reliable.

*'Language Testing' (A. Davies) - in Vol II, Nos. 3 & 4 of Language Teaching and Linguistics: Abstracts, (July and October 1978; two parts, C.U.P.)

B. RESOLVING CERTAIN DIFFICULTIES IN THE
DEVELOPMENT OF COMMUNICATIVE TESTS

The most enlightening and helpful approaches to communicative testing were those discussed with Morrow at Reading University. Keith Morrow points out a number of possible solutions to the problems involved in designing communicative tests. These are discussed fully in his article 'Communicative Language Testing - Revolution or Evolution' (1978)*. Two particular aspects of his approach are outlined below.

One of the reasons why teachers require communicative proficiency tests is to measure in a fairly standardised, accurate way what a pupil can actually achieve through language. From this, it becomes clear that our main concern is not with purely communicative competence, i.e. the degree of knowledge the pupil may have about the use of the language. We are more interested in what a pupil is actually able to do with language, i.e. his communicative performance. This in turn implies performance-based tests and in his article, Morrow develops the fairly novel idea that language performance rather than language competence is a legitimate area of concern for tests. In particular, he discusses the two main problems of extrapolation and assessment posed by this approach.

i) Extrapolation

If one assesses pupils' performance in terms of a particular task, what can one learn of their ability to perform other tasks. The starting point for such an assessment is a model for the performance of global communicative tasks showing the enabling skills which have to be mobilised to complete the tasks. Such a model implicit in J. Munby's Communicative Syllabus Design (O.U.P. 78) has been refined for assessment purposes by B.J. Carroll (1978). For instance in a test of reading, the starting point could be to ask questions such as, "What sort of reading do people do?" or "Why do people read?" These questions give a set of authentic global tasks, e.g. to skim for gist; to search for information, etc.

*Communicative Approach to Language Teaching Ed. K. Johnson,
O.U.P.

Morrow gives the following example of a global task with its enabling skills:

<u>Global Task</u>	- Search for specific information
<u>Enabling Skills</u>	- e.g. Distinguish main point from supporting details.
	- Understand text relations through grammatical cohesion devices.
	- Understand relations within sentences.
	- Understand conceptual meaning.
	- Deduce meaning of unfamiliar lexis.

The application of these skills extends beyond any one particular instance of performance. Also an analysis of the global tasks tends to show a fairly consistent set of enabling skills. Consequently, they offer a possible approach to the problem of extrapolation.

Production tests (i.e. writing and speaking) could be assessed in relation to both specified global tasks and the strategies and skills which have been used to achieve these tasks. However, with Receptive tests (i.e. listening, reading and comprehension), Morrow points out that questions focussing on specific enabling skills are needed for extrapolation to be convincing because "it may be rather more difficult to see just how the global task has been completed. For example, in a test based on the global task exemplified above it is rather difficult to assess why a particular answer has been given and to deduce the skills and strategies employed" (Morrow, 1978). This would be particularly so where the test format is one of a number of true/false questions which the candidate has to answer by searching through a text.

ii) Assessment

The two basic problems involved in assessment are "how to judge production in ways which are not purely subjective and how to assess a particular set of receptive tasks which are appropriate for different levels of language proficiency?". On this issue, Morrow suggests that the answer seems to lie in the concept of an operational scale of attainment, in which different levels of proficiency are defined in terms of a set of performance criteria, such as in the work of B.J. Carroll (1977), where "he distinguishes different performance levels by matching the learner's performance with operational specifications". Morrow summarises these specifications as follows:

C. CLOZE TESTS AND DICTATION

Cloze tests and dictation are both examples of integrative skills assessment procedures, i.e. where the receptive and productive skills (of listening/reading and speaking/writing) are combined. Oller and some of his colleagues (Streiff, Conrad, Irvue Atai, etc.) claim that both procedures test basic language processing mechanisms and sample a wide range of structural and lexical items in a meaningful context. Oller and Streiff (1975) claim that "the listener's mechanism which operates in dictation is a grammar of expectancy". Davies also notes that: "Since Cloze is also tapping the grammar of expectancy, this leads to the high correlation found between dictation and Cloze". (Davies 1978, p219).

In recent years Cloze tests have enjoyed great popularity but since the early 1970's a number of researchers have expressed doubt about whether they actually measure what they are supposed to measure. Research (Alderson et al) shows that Cloze may be more a measure of redundancy than comprehension and a better measure of the readability of the text than of reading comprehension. In 1978, Alderson came to the conclusion that what Cloze and dictation have in common is a measurement of formal linguistic skills. Morrow (1978) says that neither Cloze tests nor dictation give convincing proof of a candidate's ability to actually use the language i.e. to translate the competence (or lack of it) which he is demonstrating into actual performance "in ordinary situations" by using the language to read, write, speak or listen in ways and contexts which correspond to real life.

It is clear that both types of tests need to be used with caution. However, a visit to the Institute of Education led to an interesting discussion with Ken. Cripwell (E.F.L. lecturer and Chief Examiner, R.S.A.) on the use of Cloze tests. The main points to come out of this discussion were that Cloze procedures are only useful as a supporting technique in helping to determine the cut-off point at which children no longer require language support. Cloze procedures have one advantage in that although they are crude, teachers find them practical to administer. If Cloze procedures are used there are a number of important points to remember. Firstly children need to be at discourse level (possibly 7-8 years, for mother-tongue English children), and able to understand passages at chapter level. Therefore, a 200-250 words minimum length is essential and the passage needs a 'run-in' and a 'run-out' section. The technique also needs to be practised. Secondly it is better to use more than one passage and then let the children choose for themselves according to their own interests.

D. SUMMARY

In the discussion above, it can be seen that there is still a considerable amount of debate about the emphasis that needs to be placed on communicative assessment procedures and the feasibility of developing such tests. Two main problems in the development of these tests are extrapolation and assessment. K. Morrow in his article *Communicative Language Testing, Revolution or Evolution* (1978), points out possible solutions by seeing the problem as one of needing to assess language performance rather than language competence.

By establishing and analysing a group of authentic global tasks a fairly consistent set of enabling skills can be found, thus offering a possible solution to the extrapolation problem.

The assessment problems could be met by:

- i) The use of discrete point items in relation to the enabling skills.
- ii) The matching of learner's performance with an operational scale of attainment (B.J. Carroll, 1977).
- iii) Meetings of examiners to set up standardisation procedures.

With these kinds of insights and approaches communicative proficiency tests are being developed for English as a Foreign Language (E.F.L.) and English for Academic Purposes (E.A.P.), but not as yet for the E2L situation in schools.

It is now necessary to see how these approaches could also be applied to the E2L teaching situation in multi-cultural schools and the next section examines the practical problems involved.

PART 3

ADDRESSING THE RESEARCH FINDINGS TO THE PRACTICAL
PROBLEMS OF E2L TEACHING IN MULTI-RACIAL SCHOOLS

- A. VARIABLES
 - i) The stage variable
 - ii) The age variable
- B. AGE-GROUPS
- C. FORMATS, TECHNIQUES AND ADMINISTRATION
RELATED TO THE GENERAL AGE GROUPS OF:
 - i) infant
 - ii) junior
 - iii) secondary
- D. SUMMARY

A. VARIABLES

i) The Stage Variable

One aspect of the communicative approach to E2L testing is that there is unlikely to be a single overall test of language proficiency. Instead there will probably be different tests of proficiency at different levels with specified communicative criteria, e.g. proficiency for beginners, intermediate or second phase and advanced levels.

This is mainly because different E2L children will have been learning English as a second language for varying lengths of time and with varying degrees of success and so children of the same age may be at different stages.

Communicative testing has as its starting point the measurement of what pupils can achieve through language. Such tests should enable children to produce their own language. It is important that their level of proficiency be related to the particular stage they are at, i.e. beginner, second stage, etc.

ii) The Age Variable

When looking at the question 'Do children of different ages need a different approach or just different materials?' it soon becomes clear that the content of communicative tests also needs to be related to different age groups of children as well as different levels of proficiency. This is because there is continual perceptual and conceptual development throughout childhood and the variation between particular age-groups, is considerable. It has been shown that whilst almost all the basic linguistic apparatus of mother-tongue is acquired by the age of 5, it is evident that the learning of details, exceptions, special cases, nuances and niceties goes on for much longer - even in adulthood. Carol Chomsky's 'The Acquisition of Syntax in Children from Five to Ten' (M.I.T. Press 1969) gives some interesting examples of later language acquisition. For example, she found that older children are better able to interpret sentences in which aspects of underlying meaning are only implicitly signalled. This is illustrated by referring to the fact that many seven year olds, and some nine year olds were confused by 'ask' and 'tell'. Many could not follow an instruction beginning with "Ask/tell your friend what to feed the doll." This sentence is apparently difficult because a clause has been deleted. Children's performances were improved when the missing material was provided:

- e.g. 1 Ask your friend what to feed the doll
2 Tell your friend what to feed the doll

compared with

- 3 Ask your friend what you should feed the doll
4 Tell your friend what she should feed the doll

In the C. Chomsky study, it was found that 1 and 2 were more difficult for younger children to interpret than the extended sentences i.e. 3 and 4. It would seem then that language continues to develop at different age levels. "Throughout childhood children develop both in their ability to process language more efficiently and in their mastery of complex details of the language." (Slobin). Any communicative test for school children needs to be developed directly in relation to the language needs and situation of the children for whom it is intended. The age range in British schools can be from 4 years to 18 years. Consequently the most important general variable which will affect the content and format of the assessment procedure will be the 'age variable'. It will be of major importance in choosing which skills, situations, notions, structures, varieties and usages of language are appropriate to the language needs of the children.

On the other hand, the communicative approach as (proposed by Morrow) of setting authentic communicative performance tasks would be appropriate for all age groups, though the precise items of the 'test' would need to be different according to the age of the child.

One point worth remembering with these communicative performance tests is that once the subject content has been established according to the language needs, multi-cultural situation and age of the children, the different levels of proficiency (beginners, second stage, advanced) can be assessed by using the same materials. This could be done by setting different sets of tasks for the children to perform on the materials appropriate to a particular age, e.g. using a shopping situation, tasks might be set at three different levels:

- e.g. a) Asking for a specific item
b) Asking for items with different qualities or quantities
c) Changing defective goods

B. AGE GROUPS

As the E2L communicative tests will relate both to the ages of children and to different levels of proficiency a major problem could be the number of tests required. Given limited resources for developing such tasks, it would be useful if some age groups could be bonded together.

This will depend mainly on the developmental features of different age groups but much more research information is still necessary in this field.

Age-groups which have often been treated as homogenous, e.g. infants, juniors (lower/upper) and secondary (lower/upper) have largely been determined by administrative factors rather than precise analysis of developmental factors, e.g. a system operating first schools, middle schools and higher schools would combine age-groups in a different way from a system based on a primary/secondary division.

Any test which is to be suitable for a particular age range of children should be appropriate in terms of (1) their conceptual skills, (2) their general language skills (speaking, listening, reading, writing) and (3) their expected reading ability. Just as important the illustrations and contextualisation that a test uses also need to be relevant to the children in terms of what they would normally come across in their day-to-day experience in or out of school. Test materials must be appropriate both in conceptual and in experimental terms.

Conceptual: e.g. nursery children may need to work with three-dimensional rather than two-dimensional materials.

Experimental: e.g. test materials relating to 'Holidays by the Seaside' may not be within the experience of many Asian children.

In the time available for this research project, it has been impossible to go more deeply into the suitable age grouping of children for assessment purposes. However, it would clearly be well worth while to follow this up especially for feedback into any syllabus language development both for E2L and English mother-tongue children.

In recent years an increasing amount of material* has been produced on first-language acquisition. These materials, for example the observation technique used by J. Tough could be a useful starting point for looking at language development which is one of the main developmental features to be considered when grouping children of different ages.

* For example: the work of Roger Brown, Bellugi, Carroll, Menyuk, etc. (See bibliography and 'Seven Questions about Language Development' by Dan I. Slobin). There has also been a lot of interesting work done on second language acquisition, see 'Error Analysis - Perspectives on Second Language Acquisition', Edited by J. Richards, Longman's 1974, and E. Ingram and R. Clark in the Edinburgh Course Book, Vol 2, O.U.P. 1975.

C. RELATING THE FORMATS, TECHNIQUES AND ADMINISTRATION
OF TESTS TO THE GENERAL AGE-GROUPS - INFANT, JUNIOR
AND SECONDARY

i) Infant

In order to interpret any child's performance on a task one must be reasonably sure of how the child comprehends the task. The language of instruction must be congruent with the way in which a child conceives of a task. In such cases tasks are easier to carry out promptly and correctly than when instructions are at variance with the child's conception.

Example: A young child of 4 or 5 years of age is playing with coloured bricks. An instruction such as "Can you build it so that the red block is on top of/under the blue block," is easier if the child is holding the red block rather than the blue one. This is because the child makes a link between the subject of the sentence and the block under his control.

Young children need experience of conceptual development for learning although they can also learn away from concrete situations but this is done mainly by reflection about their own actual experience. Motivation, through play, exploration of materials and observation of the social and physical world, builds up children's understanding and develops their skills. Therefore, authentic tasks relating to oral and listening skills are best set up on the basis of drawing on the children's experience and using the above learning situations. The main point is that teachers and assessors should be aware of the language they are using to set up tasks.

Assessment will probably be through building up individual profiles based on working through a specified set of tasks which reflect the children's language needs.

The work done by J. Tough (Communication Skills in Early Childhood - Schools Council Project), could be utilised to examine the purposes for which children use language in the classroom. According to this work, young children use language most frequently for the following purposes:

- For gaining attention and approval.
- For identifying objects and actions.
- For stating what they have or what they want to have.
- For telling other people what they want them to do.
- For asking permission to do or take things.
- For claiming their possessions.
- For defending themselves.
- For asking where things are and following directions to find them.
- For making a running commentary on their activities.
- For anticipating a way of working.
- For directing others and following the directions of others.
- For exploring new situations, for finding out about people.
- For imagination in play.

Joan Tough categorises these uses into four main headings:

- 1) Self-maintaining
- 2) Directing
- 3) Reporting
- 4) Reasoning

Children need language for:

- 1) Self-help, i.e. 1, 2, 3, above
- 2) Joining in, i.e. strategies of reporting and directing, strategies for anticipating and prediction.
- 3) Finding out and enquiring
- 4) Extending learning

An interesting ability that young have is that of being able to use different varieties of language. Children even at 4 years of age are seen to be capable of modifying their language according to the situation. This has been particularly noted with young children communicating with younger brothers and sisters. For example, they would probably use more repetition and more simple declarative sentences. (Slobin, 1969).

Consequently, an overall assessment of proficiency would need to contain performance tasks which can also allow children the opportunity to express to some degree this kind of flexibility in language use. One extremely useful technique for this would possibly be through the use of dialogues and puppets.

It should be clear from these points, that the location of assessment procedures ought to be within the context of everyday activities and avoid the abstract forms which many tests take.

ii) Junior

Language provides the most important medium through which thinking can be expressed and it appears that children's use of language also affects the way in which they will think and the kind of interpretation they will make of their experiences.

If we consider the kind of learning that is expected of children in school, we can see that this becomes increasingly dependent on being able to use past experience. In act 'Reporting' seems to provide the basis from which all other language uses can develop, such as 'logical thinking' and 'problem solving' as well as for 'imaginative use'. This is clearly essential for example in topic work following historical themes, primary science projects etc.

Nicholas Hawkes at the York University 'Language Materials Development Unit' has been one of the co-developers of the Kaleidescope materials for teaching junior-age children E.F.L. in Western European schools. A large amount of unpublished assessment materials were produced by Nicholas Hawkes for the Kaleidescope course.

These techniques include:

Pictures of situations with introductory dialogue. Children complete the dialogue in empty speech balloons showing what the person is saying.

Pictures again using empty speech balloons, showing someone who is in need of advice, "What should be said/suggested/advised?"

Identifying social contexts by choosing correct pictures.

Showing acceptability of sentences in relation to situations (via pictures) - YES/NO.

Logical elimination - given clues (tape/text) and eliminate a picture in turn.

Using plans@diagrams - e.g. plan of school, "Where is?". Take someone on a visit round school.

"Shake the dice and ask" games

Challenges - "Can you" (pictures)

Discussion on a timetable.

Essence game i.e. looking for common features from different pictures.

Which are you? Identifying characters by question and answer.

(Kaleidescope - unpublished test materials - N. Hawkes).

Some of these techniques have been used to assess discrete point structural items but they also have clear potential for being used in a communicative way.

iii) Secondary

The difficulties of developing communicative proficiency tests for secondary children tend to be less than those for infant and junior children. At school-leaving age, the specific language needs of children are relatively easier to define. A lot of work has already been done in this area and apart from the cultural aspects of tests the difference between an E.F.L. communicative test and an E2L communicative test is smaller.

As noted earlier a number of C.S.E. papers have been developed for school leavers (although on closer examination, they do not always assess what they claim to be assessing, i.e. communicative competence.)

There is more difficulty in assessing younger secondary school children as no suitable assessment procedures really exist for this group. Many of the techniques and ideas proposed for junior children would be appropriate for lower secondary children, although language tasks may not need to be as practically based. All four general language skills of speaking, listening, reading and writing, could be utilised in overall assessment although different emphasis would be placed on them. Literacy becomes particularly important in the secondary school but teachers of E2L will still be conscious of the need to maintain and develop oral work on the basis that for many E2L learners the receptive skills (listening and Reading) may well be more highly developed than the productive skills (speaking and writing).

When preparing their pilot study for 'English as a Second Language' (C.S.E.), the North West Regional Board pinpointed five basic core functions of language which they consider to be typical of the widest range of expository English occurring in most school subjects and in life outside school.

The listed core functions are:

- 1) Description: specific and general
- 2) Narrative: important in writing up reports and in the historical and development aspects of most school subjects, especially useful when relating items in a sequence.
- 3) Instruction: how to make things, how to do things, how to get somewhere as well as giving tips for safety, taking care of appliances and so on.
- 4) Process: this is a kind of description but one which has to follow a particular sequence relating to the order in which steps are carried out. It is essentially an analytical process.
- 5) Narrative padding: background description within a narrative sequence.

This study shows how basic language items, typical of each 'core function' were isolated by focussing on a correlation of verb forms and sentence pattern so producing a workable description of the language used in each area. A full report of how the pilot study was developed is contained in the pilot study booklet.

D. SUMMARY

During the discussion in this section, brief outlines of J. Tough's classification of language functions and uses and the North West Regional Board's language functions have been outlined. These two approaches can be used to illustrate some of the language requirements demanded of children of different ages, although they do overlap. If these approaches are to be utilised for building up communicative proficiency tests, we need to re-emphasise the different aspects of performance-based tests. We need to ask again why children read, write, speak and listen and to ascertain what enabling skills they actually need to enable them to use language for these particular purposes.

PART 4

GUIDELINES FOR DRAWING UP ASSESSMENT PROCEDURES
FOR COMMUNICATIVE PROFICIENCY TESTS

- A. GUIDELINES 1 - 7
- B. GENERAL MODEL OF DESIGN FOR
COMMUNICATIVE PERFORMANCE TESTS
(Morrow)

GUIDELINES FOR DRAWING UP ASSESSMENT PROCEDURES
FOR COMMUNICATIVE PROFICIENCY TESTS

There are a variety of problems involved in the development of communicative proficiency tests. These include working out the language needs of children in terms of global tasks and enabling skills; the age relevance and the authenticity of materials; extrapolation and the actual form of assessment. Nevertheless these problems are not insoluble and the following guidelines may be helpful.

A. GUIDELINES

Guideline 1: Analysing Language Needs

In order to develop effective language tests a detailed specification of the communication needs of the potential users of the language is necessary. In other words, the communicative testing framework needs to be based on an examination of the purposes for which children use language e.g. the language which enables them to participate in the kinds of learning expected in the classroom.

In order to develop working lists of comprehension tasks, enabling skills and language functions in the classroom, language profiles could be built up for a variety of children from different schools, classes, age-groups and subject specialisation. This could be achieved through some of the following techniques and sources of information.

- a) Observation and analysis of classroom language and its relationship to the curriculum.
- b) Teacher study groups (e.g. N.A.M.E. or N.A.T.E.)
- c) Publications of a general linguistic nature. CARROL, MORROW, MUNBY etc. (See Bibliography).
- d) Publications specific to the language of school children. HAWKES, TOUGH, VAN EK etc.
- e) Other relevant sources of information e.g. Assessment of Performance Unit, National Foundation for Educational Research, Centre for Urban Educational Studies (I.L.E.A.)

Guideline 2: Age Groups for tests

Discussion needs to take place on the most appropriate age-grouping of children bearing in mind both developmental and organisational constraints. (See Part 3, section B 'Age Groups')

Guideline 3: Language banks of 'operations' and enabling skills

Following Morrow's approach the main starting point is identifying and making working lists of the language 'operations' or tasks, that involved in comprehension, oral interaction and writing. Receptive tests (listening and reading) also require working lists of the enabling skills involved in each task, so that greater generalisation can be made about a pupil's language ability from his performance on one particular task.

An important but problematical area in testing spoken language is the analysis of oral interaction in operational terms. It is essential that oral interaction tests are operational ones.

Three general purposes can be identified in talking and, most interchanges involve a combination of these three elements. Either we are trying to get information or give it; or we are involved in changing people's behaviour; or we are trying to express, ascertain or interpret attitudes, emotions and relationships. We do this using three basic strategies, either we 'say' and convey information; or, we 'ask' and elicit information; or we 'interact' with someone in a conversation. Morrow suggests that if we relate these three strategies to the general communicative purposes in speaking, we can establish an operational basis for test procedures.

e.g.

SAY (convey)	Express thanks/needs etc. Narrate etc.
ASK (elicit)	Help information etc.
INTERACT	Say and ask specific interaction skills e.g. Carroll's Language Skills specification i.e. Initiating discourse Maintaining discourse Terminating discourse

Guideline 4: Choosing which 'operations' and 'enabling skills' to use

Decisions need to be made as to which operations and enabling skills are to be tested, appropriate to proficiency level and age of the children. The limiting developmental factor of age should be considered here but decisions will also have to be based on feedback from piloting stage of the assessment procedures. (See Morrow's example of a global task with its enabling skills p.1)

Guideline 5: Authentic materials

The language tasks set should be authentic ones, i.e. the type of task that people actually carry out in using language. This should also apply to the materials used, i.e. use of texts - text books, work cards, notices, etc. T.V. and radio programmes, talks, etc. Collections of suitable materials related to specific language tasks will be necessary.

Guideline 6: Administration

Comprehension Tests and Writing Tests can be administered to a number of children at a time but oral interaction tests could also involve more than one pupil (i.e. pupil to pupil), or more than one teacher (i.e. pupil to teacher with another teacher-assessor present).

Guideline 7: Assessment

According to Davies (1978), "Morrow argues that communicative testing can employ either discrete point items or integrative items." Examples of discrete point items would be manipulative ones (e.g. multiple-choice questions based on a given stimulus. Discrete point items may also be productive (e.g. a request to produce appropriate language from an array of choices in a given situation). Examples of integrative items would be: answering questions freely, asking questions freely, participating in interaction, providing a narrative, etc. Morrow describes an interaction participation test in which the testee's part in a dialogue is blank; the learner therefore has to find fillers for the blanks in the dialogue.

The blanks may be of varying lengths ranging from single words to complete utterances. It should be noted that the smaller the blank, the more the test becomes one of grammatical rather than communicative competence.

As the emphasis in communicative tests is to see how a child can use language, plenty of opportunity must be given to allow the child to use his/her own language. Productive tests will involve a lot of role playing for pupils either projecting themselves into different situations or different characters. Puppets could be extremely useful here, especially with young children. B.J. Carroll's work could be refined to work out an operational scale of attainment related to a set of performance criteria (see Appendix II). Standardisation procedures would be essential for the preparation and marking of productive tests.

These then are some of the basic guidelines that should be followed and kept in mind when drawing up assessment procedures.

B. GENERAL MODEL OF DESIGN FOR COMMUNICATIVE PERFORMANCE TESTS
(Based on the work of K. Morrow and colleagues).

- 1) Establish the types of operations to be tested in each mode.
- 2) Decide which of these operations are to be tested at each level and in what proportion.
- 3) (Optional: Receptive tests only)
Decide which enabling skills are to be tested at each level and in what proportions.
- 4) Examine the Event/Activity inventory to establish the areas in which the operation might be authentically carried out.
- 5) Receptive: Locate authentic text from relevant area.
Productive: Set up authentic task within relevant area.
- 6) Decide on format for 'questions' and assessment procedures.

Some form of qualitative assessment procedure, other than discrete item testing, will be necessary, particularly for speaking and writing tests. This requires skilled assessment rather than simply adding up correct answers.

PART 5

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS AND RECOMMENDATIONS

The teacher of E2L is faced with many decisions about what to teach and how to teach it. The case for E2L support is clearest with children who have no English or are in the first stage of E2L work. Specialist teachers of E2L need to be accountable for their involvement if it is to be argued that such involvement is far more important than 'letting children pick it up as they go along'. This accountability extends beyond the 'first stage' of second language work.

I have indicated that traditional tests have been limited to discrete item tests emphasising vocabulary, spelling, meanings and structures or correct sentence patterns. The main thrust of this research however, has been to identify the target for our pupils as communicative competence and performance. E2L communicative performance tests could help teachers to be more effective in their E2L teaching and in monitoring the language development of their pupils. Communicative tests could also assist teachers in making decisions as to which children require or alternatively no longer need, specialised E2L help. Given this, and the fact that it is possible to develop standardised communicative performance tests along the lines suggested in this report, great benefit would be derived from devising such tests.

I would therefore recommend that a working party* be set up to: *develop:*

- 1a) Language profiles of the communicative needs of children in school (and in the community at large).
- b) Working Lists of 'operations' or 'tasks' for comprehension and oral interaction tests (relevant to age groups).
- c) Working Lists of 'enabling skills' for comprehension tests (relevant to age groups).
- d) Language banks of authentic materials for specific assessment purposes.
- 2) Produce formats and items for the tests (relevant to age groups and proficiency levels).
- 3) Devise a piloting scheme for the specified assessment procedure including the development of an appropriate scoring system.

* Ideally such a working party would consist of E2L teachers and also other members of staff involved in mainstream teaching.

CONTENT OF THE TESTS AND SKILLS TESTED

Difficulty Levels		Skill Areas			
		Listening	Speaking	Reading	Writing
Level One	Name of test Skills tested	Listening One (33 items) Recognition and comprehension of spoken words (in simple context)	Speaking One (25 items) Production of spoken words	Reading One (33 items) Recognition and comprehension of written words (in simple context)	Writing One (25 items) Production of written words
Level Two	Name of test Skills tested	Listening Two (28 items) Comprehension, in terms of both vocabulary and grammar, of spoken sentences	Speaking Two In three (5 items) parts (10 items) (10 items) Production of spoken phrases, sentences, and/or questions	Reading Two (28 items) Comprehension, in terms of both vocabulary and grammar, of written sentences	Writing Two In two (10 items) parts (10 items) Production of written phrases and/or sentences
Level Three Part One	Name of test Skills tested	Listening Three part one (14 items) Comprehension of a spoken word, presented in context, and choice of synonym from list provided	Speaking Three part one (10 items) Production of more complex spoken sentences	Reading Three part one (15 items) Comprehension of a written word, presented in context, and choice of synonym from list provided	Writing Three part one (10 items) Production of more complex written sentences
Level Three Part Two	Name of test Skills tested	Listening Three part two (16 items) Skill in following and understanding an extended piece of speech	Speaking Three part two (1 task) Production of an extended piece of descriptive speech	Reading Three part two (15 items) Skill in following and understanding an extended piece of writing	Writing Three part two (1 task) Production of an extended piece of writing

APPENDIX II

Matrices of Performance Criteria - matching a learner's performance with an optional scale of attainment.

B.J. CARROLL (1977)

Carroll has constructed three matrices which attempt to classify language performance at three levels.

- a) Basic
- b) Intermediate
- c) Advanced

Language performance is divided into four subdivisions:

A. Listening B. Speaking C. Reading D. Writing

These are set against a range of performance criteria i.e.

- 1) Size
- 2) Complexity
- 3) Range
- 4) Speech
- 5) Flexibility
- 6) Accuracy
- 7) Appropriacy
- 8) Independence
- 9) Repetition
- 10) Hesitation

Some examples taken at random from across the three matrices will illustrate the approach.

Basic Skills

Skill-'speaking' Criterion-'range'

"Range of topic and skill restricted to lesson/syllabus experience and familiar topics. Limited range of tones and functions."

Intermediate

Skill-'listening' Criterion-'complexity'

"Grasps two or three major points with their subsidiary themes. Can cope with some noise and irrelevancy in detail. Does not have to see speaker. Copes with two or three topics."

Advanced

Skill-'writing' Criterion-'appropriacy'

"Uses style and register suitable for purpose, uses slang and attitude tone markers effectively."

'An English Language Testing Service' - B. J. Carroll
(1978) English Language Division, British Council.

Details of the participant: e.g. identity, language, age.

Purpose of Study: establishing the type of English and the purpose for its use.

Settings for English: including both physical and psychosocial settings, e.g. school, classroom, playground, science rooms, etc.

Interactions involved: identifying those with whom the participant will communicate, in English, his position, role relationships and social relationships, e.g. pupil/teacher, pupil-pupil, pupil-dinnerlady, pupil-senior/junior, etc.

Instrumentality: the main activity areas-receptive/productive, spoken/written; the channels, e.g. face-to-face, print or T.V.

Dialects of English: that the children will encounter or need to understand, e.g. need to understand British Standard English dialect and produce acceptable regional version of standard English accent.

Proficiency Target levels: expressed on a scale and related to the dimensions of "test size, complexity, range and delicacy, and the speed and flexibility of handling conditions related to tolerances of error, style, reference, repetition and hesitation.

Communicative Events and Activities: the description of what participants have to do, such as participating in a lesson (event), and the parts of those events that assist skill selection later, such as "putting forward one's point of view" (activity).

Attitudinal Tones: concerning how an activity is enacted derived from an index of attitudinal tones - sets of antonymous continua, such as 'formal-informal'.

Language Skills: a taxonomy of 54 skill categories, with their component skills.

Example Item 24 - Understanding Conceptual Meaning.

- (i) Quantity
- (ii) Definiteness
- (iii) Comparison
- (iv) Time
- (v) Location
- (vi) Means
- (vii) Cause etc.

Micro-function: units of meaning within one level. e.g. "putting forward one's point of view" using micro-functions such as persuasion, advising, invitation, etc.

GLOSSARY

N.B. Some of the linguistic terms included in the text and in this glossary have a different meaning in everyday usage.

<u>Appropriacy:</u>	Correct <u>use</u> of language within a given situation and context - e.g. "Give me that ball!" This sentence is correct grammatically. If a child uses it with one of his peers it would also be an appropriate/correct use of the language but it would be inappropriate to request the ball from his teacher in this manner. "Can I have the ball please?" would be appropriate here.
<u>Cloze Test:</u>	A procedure of missing out words in a passage of prose, usually at fixed intervals. The pupil is required to fill in the appropriate words. The result is supposed to give the tester an understanding of the pupil's reading comprehension.
<u>Concept:</u>	The basic meaning(s) of a lexical item or a structure, e.g. a particular use of a tense.
<u>Context:</u>	Often used to refer to the significant linguistic features surrounding an utterance; ' <u>Situation</u> ' would then refer to the non-linguistic features surrounding that utterance.
<u>Contextualise:</u>	Make sentences relevant by referring to a specific person, object or situation.
<u>Communicative Competence:</u>	The ability to combine the elements of a language - e.g. structures, vocabulary, etc. in new and appropriate ways to meet the linguistic demands of the situation in which a learner wishes to use the language.
<u>Communication Goal/Function:</u>	What the speaker intends to convey as opposed to how he does it, 'asking a question' would have a communicative function 'forming an interrogative' would not.
<u>Criterion - Reference Testing:</u>	Testing which relates a pupil's performance to a particular linguistic task (criterion) - giving information about an individual's progress or ability in relation to specific tasks or objectives. This is unlike ' <u>norm-reference testing</u> ' which relates pupils to their peer groups and gives rank order information

<u>Global Tasks:</u>	Where a language skill (e.g. reading) is used to complete a particular overall operation or task, for example - skimming for gist, searching for specific information would be global or operational tasks, requiring specific language skills.
<u>Grammatical/Formal Linguistic Competence:</u>	(cf. Chomsky) - Ability to formulate only the grammatical sentences in a language.
<u>Grammatical Form:</u>	Normally used as a contrast to semantic (meaning) aspects, i.e. concerned with the form rather than the meaning.
<u>Integrative Tests:</u>	Where the receptive and productive skills of listening/reading and speaking/writing are combined in one test item.
<u>Interaction:</u>	Features which occur during e.g. conversation. As such, its use would normally be restricted to spoken language.
<u>Interlanguage:</u>	Is a term used by Selinker to suggest the intermediate stages between the learner's mother language and the target language.
<u>Lexical Item:</u>	An item of vocabulary; this can be a combination of more than one word, e.g. 'door stop', 'make-up for', but not a structural item - e.g. pronouns, prepositions.
<u>Reading Extensive:</u>	Reading for general comprehension without concentration on detail.
<u>Reading Intensive:</u>	Reading with the aim of complete comprehension of all features.
<u>Register:</u>	Usually a functional variety of language, e.g. journalese, 'hippy', legal; the items explained in this glossary are from the 'register' of language teaching. Often used to include 'style'.
<u>Reliability:</u> <u>(of a test)</u>	A test is reliable if it produces the same result twice with no learning or teaching between the two occasions. Reliability can be affected by: - scope of test items - conditions of administration - clarity of instructions - personal factors: motivation, illness, etc. - reliability of markers - objective or subjective test N.B. Reliable tests are not necessarily <u>valid</u> (see under validity).

Discrete Point and Discrete Items:

By 'discrete point' is meant two things (Aitken, 1976):

- 1) Analysis of the language on some taxonomic basis, e.g. phonology, semantics or vocabulary, and then subdividing, e.g. phonology into phonemes, stress, intonation, rhythm. In other words, the analysis is in the terms of structural linguistics. The 'discreteness' of items is a quality of their being categorised within a taxonomic system which is set up so as to contain discrete items.
- 2) The second meaning of discrete point: since the linguistic analysis consists of linguistic items which do not overlap with one another, then the task presented to the students is to distinguish between one discrete item and another, i.e. between items which differ from one another by only one feature e.g. multiple choice items. Typical discrete point (multiple choice) item:

I want ... (a) going
(b) go
(c) to go
(d) to going home now.

(Survey Article - Language Testing
by Alan Davies.)

Discourse Function:

Communicative function of a piece of language within the context of a discussion, conversation, etc.

'Formal' Linguistic Explanation:

The expression a 'formal' linguistic explanation of language basically means an explanation relating to the grammatical forms and structures of language rather than relating to the semantics of language.

Formally:

Normally used as a contrast to 'semantically', i.e. practising or demonstrating the 'form' rather than the 'meaning'.

Functional Approach:

An approach to teaching or learning of a language which concentrates more on effective communication, as opposed to a formal structural approach.

- Structuralist Approach: This approach starts from the implicit assumption that mastery of a language is possible by isolated mastery of its separate parts. The approach is based on structural linguistics and behaviour learning theory.
- Style: Usually a social variety of language, e.g. formal, informal, casual.
- Taxonomic Analysis: This involves the categorising of language into various elements and levels, e.g. phonology, grammar, semantics and vocabulary. Phonology for example, can be divided into: phonemics, stress, intonation and rhythm.
- Transitional Competence: Is basically a learner's underlying knowledge of the language to date.
- Utterance: Any single speech act, i.e. single words and incomplete sentences as well as 'full' sentences.
- Validity (of a test): The test measures what it is intended to measure.
- (a) Face validity: e.g. culturally relevant to pupils.
 - (b) Content validity: relates to what was actually taught or what really needs to be assessed and not to what is only easiest to assess.
 - (c) Construct validity: this assumes a particular learning theory and emphasises features of this particular theory tending to ignore aspects, e.g. if 'structuralist', the test tends to emphasise the structures of English with rather less stress on situational meaning.
 - (d) Predictive validity: empirical or statistical validity gained by comparing test results with other forms of measure, e.g. another test, teachers' assessment, etc.

BIBLIOGRAPHY

BOOKS

- ALDERSON, J.C. (1978) A study of the Cloze procedure with native and non-native speakers of English - University of Edinburgh Ph.D. thesis.
- CHOMSKY, C. (1969) The Acquisition of Syntax in Children from Five to Ten, M.I.T. Press.
- DAVIES, A. (1968) Language Testing Symposium, O.U.P.
(Editor) Davies & E. Ingram, 'Attainment and Diagnostic Testing'.
- EK, J. VAN (1976) The Threshold for Modern Language Learning in Schools for Council of Europe, Longmans.
- HAYNES, J. (1971) Educational Assessment of Immigrant Pupils. N.F.E.R.
- HEATON, J.R. (1975) Writing English Language Test. Longman.
- HEGARTY, S. (1978) Able to Learn? The Pursuit of
& LUCAS Culture and Fair Assessment, N.F.E.R.
- INGRAM, E. (1974) Language Testing p1313-343, Edinburgh Course, Vol. 3, O.U.P.
- JAKOBOVITS, L.A. (1970) Foreign Language Learning Chapters 3 & 4. Problems in assessing Language Proficiency, p149-221. Newbury House Publishers, massachusetts.
- LADO, R. (1961) Language Testing: the construction and use of foreign language tests. Longman.
- MACKEY, W.F. (1965) The measurement of Language Learning p403-416. Language Teaching Analysis. Longmans.
- MUNBY, J. (1978) Communicative Syllabus Design. O.U.P.
- RICHARDS, J.C. (1974) Error Analysis: Perspectives on Second Language Acquisition. Longman.
- ROSEN, C. and R. (1973) The Language of Primary School Children. Penguin.

- TOUGH, J. (1976) Listening to Children Talking.
Ward Lock Educational.
- TOUGH, J. (1977) Talking and Learning.
Ward Lock Educational.
- TOUGH, J. (1977) Development of Meaning:
A Study of Children's Use of Language.
Allen & Unwin.
- VALLETTE, R. (1977) Modern Language Testing:
A Handbook. Harcourt, Brace, World Inc.
- WILKINS, D.A. (1976) Notional Syllabuses, O.U.P.
- WILKINSON, A. (1971) The Foundation of Language, O.U.P.

ARTICLES

- BELLUGI, U. (1968) Linguistic Mechanisms Underlying
Child Speech, H. Zale, Editor.
Proceedings of the Conference on Language
& Language Behaviour, Appleton Century -
Crofts.
- BROWN, R. (1973) Development of the First
Language in the Human Species.
American Psychologist - Vol. 28, No. 2
- BRUMFIT, C. (1978) Book Review: Notional Syllabuses
D.A. WILKINS, British Association for
Applied Linguistics.
- BRUMFIT, C. (1978) Book Review. Notional Syllabus
D.A. WILKINS, E.L.T.J. 33/1.
- CARROLL, B.J. (1978) An English Language Testing
Service. Specifications. London,
British Council, Mimeo.
- CORDER, S.P. (1967) The Significance of Learner's
Errors IRAL, Vol. V/4 - in Error Analysis:
Editor, J.C. Richards.
- DAVIES, A. (1964) Language Proficiency Testing:
Report on E.P.T.B. Testing Project.
- DAVIES, A. (1978) Language Testing: Survey Article
Language Teaching Linguistics: Abstracts
Vol. II, No. 3 and 4, C.U.P.
- SPOLSKY, B; MURPHY, P; HOLM, W; & FERRELL, A.
(1972) Three functional Tests of Oral Proficiency.
TESOL, Quarterly, 6,3. pps. 221-235.

TESTS (REFERENCES)

E.T.I.C. Examination and Tests in English for
Speakers of other languages.
Information Guide 6.
December, 1976.
The British Council.

REA, Pauline Language Testing: A Selected
Bibliography.

C.S.E. Examining Boards:-

Associated Lancashire Schools E.B.
Mode III

East Anglian E.B.
Mode III

East Midlands Regional E.B.
Mode III

North West Regional E.B.
Mode I

West Midlands E.B.
Mode I

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere thanks to the numerous people whom I visited and who generously gave me their time and the benefit of their experience and work in this field.

I am also most grateful for the support given to me by my colleagues in the Minority Group Support Service, the teachers who took part in the assessment workshop sessions, and especially the M.G.S.S. team leaders who worked through the draft materials.

I would also like to express my special thanks to Keith Morrow without whom this paper could probably not have been written and whose inspiring insights into the problems of communicative assessment have allowed the hope of developing communicative tests to become a definite possibility.

I acknowledge also the help and advice given by Mr. M. Feeley, Adviser for Multi-Cultural Education, and Coventry Local Education Authority for giving me the opportunity to develop this study.

In addition to the resources available at the Minority Group Support Service and Elm Bank Teachers' Centre, Coventry, the following visits were of considerable help in the preparation of this report.

1. LONDON

- (a) Institute of Education Library, London University.
- (b) Dillons University Bookshop.
O.U.P. Charing Cross Road,
Foyles Educational, Marble Arch.
- (c) Centre for Information on Language
Teaching and Research (C.I.L.T.)
- (d) Brendon Carroll, English Language Consultation
Department of British Council,
20 Charlton House Terrace.
- (e) Ken Cripwell,
Institute of Education.
- (f) Monica Vincent, Institute of Education.
- (g) English Teaching Information Centre (E.T.I.C.) Library.

2. READING

Keith Morrow,
Centre of Applied Language Teaching,
Reading University.

3. EDINBURGH

- (a) S.P.H. Corder)
- (b) A. Davies) Edinburgh University
- (c) A. Howatt)

4. YORK

N. Hawkes,
York Language Materials Development Unit.

5. NOTTINGHAM

- (a) J. Chandlin
- (b) Andrew Craig - Aboretum Language Centre.

GERALDINE A. MARSHALL

Produced by:

The Minority Group Support Service,
Southfields,
South Street,
Coventry.
Cv1 5EJ

Tel: (0203) 26888

Text: Geraldine Marshall
Typing: Minority Group Support Service
Printing: Southfields Printing

COVENTRY EDUCATION AUTHORITY
Director of Education - Mr. R. Aitken